JURNAL RISET INFORMATIKA **P-ISSN: 2656-1743 |E-ISSN: 2656-1735**
Vol. 5, No. 4. September 2023 DOI: https://doi.org/10.34288/jri.v5i4.244
Accredited rank 4 (SINTA 4), excerpts from the decision of the DITJEN DIKTIRISTEK No. 230/E/KPT/2023

Sentiment Analysis of Telemedicine Applications on Twitter Using Lexicon-Based and Naive Bayes Classifier Methods

Arid Hasan[1], Yudhi Raymond Ramadhan[2], Minarto[3]

[1*), 2,3] Teknik Informatika
Sekolah Tinggi Teknologi Wastukancana
Purwakarta, Jawa Barat, Indonesia
[1*)] aridhasan06@wastukancana.ac.id , [2,*]yudhi.raymond@wastukancana.ac.id,
minarto@wastukancana.ac.id

(*) Corresponding Author

Abstract

Since the onset of the COVID-19 pandemic in Indonesia, many people have turned to telemedicine programs as an alternative to minimize social interactions, opting for consultations from the safety of their homes using smartphones and internet connectivity. Given the necessity for physical distancing and avoiding crowded places, these applications have become indispensable substitutes for in-person medical consultations. Numerous apps facilitating access to healthcare services have been introduced in Indonesia, ranging from business startups to initiatives by the Ministry of Health. Telemedicine can potentially revolutionize healthcare in Indonesia, addressing critical health challenges. A significant issue within Indonesia's healthcare system is the scarcity of doctors and their uneven distribution. With only four doctors per 10,000 people, this figure falls far below the WHO guideline of 10 doctors per 1,000. Sentiment analysis of these applications was conducted to evaluate how telemedicine applications meet public needs and offer an alternative solution. Lexicon-based and naive Bayes methods were employed to classify tweet data into positive, neutral, and negative sentiments. The results revealed 908 positive tweets, 172 negative tweets, and 168 neutral tweets, indicating predominantly positive public perceptions of telemedicine applications. The naive Bayes classifier exhibited a 74% accuracy rate, with a precision of 98% and a recall of 86%. These findings underscore the positive impact and acceptance of telemedicine applications among the Indonesian populace, emphasizing their significance in augmenting the nation's healthcare landscape.

Keywords: Sentiment Analysis; Telemedicine Applications; Naive Bayes Classifier; Lexicon Based

Abstrak

Sejak dimulainya pandemi Covid-19 di Indonesia, banyak orang telah beralih ke program telemedicine sebagai alternatif untuk meminimalkan interaksi sosial, memilih untuk berkonsultasi dari rumah dengan menggunakan ponsel pintar dan konektivitas internet. Mengingat perlunya menjaga jarak dan menghindari tempat keramaian, aplikasi-aplikasi ini telah menjadi pengganti yang tak tergantikan untuk konsultasi medis secara langsung. Berbagai aplikasi yang memfasilitasi akses ke layanan kesehatan telah diperkenalkan di Indonesia, mulai dari perusahaan rintisan hingga inisiatif dari Kementerian Kesehatan. Telemedicine memiliki potensi untuk merevolusi layanan kesehatan di Indonesia, mengatasi tantangan kesehatan yang kritis. Masalah utama dalam sistem kesehatan di Indonesia adalah kelangkaan dokter dan penyebarannya yang tidak merata. Dengan hanya 4 dokter per 10.000 orang, angka ini jauh di bawah pedoman WHO yaitu 10 dokter per 1.000 orang. Untuk mengevaluasi sejauh mana aplikasi telemedicine memenuhi kebutuhan masyarakat dan menawarkan solusi alternatif, analisis sentimen terhadap aplikasi-aplikasi tersebut dilakukan. Metode berbasis leksikon dan naive Bayes digunakan untuk mengklasifikasikan data tweet ke dalam sentimen positif, netral, dan negatif. Hasilnya menunjukkan 908 tweet positif, 172 tweet negatif, dan 168 tweet netral, yang mengindikasikan persepsi publik yang sebagian besar positif terhadap aplikasi telemedicine. Pengklasifikasi naif Bayes menunjukkan tingkat akurasi 74%, dengan presisi 98% dan recall 86%. Temuan ini menggarisbawahi dampak positif dan penerimaan aplikasi telemedicine di kalangan masyarakat Indonesia, yang menekankan pentingnya aplikasi ini dalam meningkatkan lanskap layanan kesehatan di Indonesia.

Kata kunci: Analisis Sentimen; Aplikasi Telemedicine; Naive Bayes Classifier; Lexicon Based

INTRODUCTION

So rapid is the growth of information technology at this time that it provides very significant changes in various aspects of life. Information technology shares many utilities for humans in each field to solve different problems. In the current era of technological development that greatly supports human needs, technology plays a significant role in human life, including the world of health.

Since the COVID-19 pandemic hit Indonesia, many Indonesians have used telemedicine programs as an alternative to avoid involvement with many people. This can be solved at home using smartphones and internet networks(Yuniar et al., 2022). Various parties from the business community have released applications supporting health service access in Indonesia. Startup to the Ministry of Health. The telemedicine application provides multiple features of medical consultation, lab checking, and online drug ordering. In totality, the available features have explored some of the services available in hospitals(CIKANIA, 2021), and some health service applications are growing and widely used in the country. (annur, 2022)

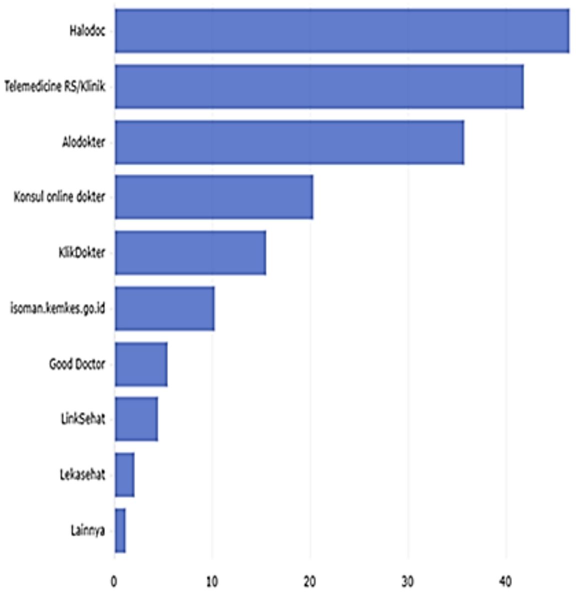

Figure 1. 1 Widely Used Application

Telemedicine is a service health professionals provide through information and communication technology, including exchanging information (Riyanto, 2021). Telemedicine has the potential to overcome various health service problems and transform the health of the Indonesian people. Doctor shortage and uneven distribution are significant problems facing Indonesia's healthcare system. Only four doctors are present for every 10,000 people, far below the WHO recommendation of 10 doctors for every 10,000 or one doctor for every 1,000 people in each country. After Cambodia, Indonesia has the lowest doctor-to-population ratio in Southeast Asia. The three ASEAN countries with the highest physician-to-population ratio are Singapore (2.3 per 1,000 people), Brunei Darussalam (1.8 per 1,000 people), and Malaysia (1.5 per 1,000 people).

Meanwhile, the distribution of doctors is uneven. According to data from the Central Statistics Agency (BPS), in 2019, There were about 11. Three hundred sixty-five practicing doctors in Jakarta, 10. Eight hundred two in East Java, 9. 747 in Central Java, 8. 771 in West Java, and 3. 126 in Banten. After that, there was a spread to Bali, South Sulawesi, Yogyakarta, Aceh, and Riau.

In contrast, the five regions with the lowest number of doctors are Gorontalo (383 doctors), North Kalimantan (349 doctors), North Maluku (324 doctors), West Sulawesi (308 doctors), and West Papua (302 doctors). Data shows more than half of medical professionals are in Java. The presence of telemedicine provides various facilities for residents, first for people in areas with minimal health services. (Kemkes, 2021). From some of the problems above, it is necessary to conduct a sentiment analysis of the Telemedicine application, whether it has met the community's needs and is an alternative solution for the community.

Sentiment analysis was conducted on social media Twitter. Another benefit of Twitter is that all posts are accessible to other users, making it one of the social media platforms often used by people to voice their thoughts. Twitter has 126 million daily users(Mahendrajaya et al., 2019).

This research uses the naïve Bayes classifier algorithm to classify positive, neutral, and antagonistic classes. The naïve Bayes classifier algorithm was chosen referring to the study tried by Solehudin Basyrah, which showed that the naïve Bayes classifier algorithm has a considerable accuracy value of above 90%. (Salehudin Basryah et al., 2021), The following refers to a survey conducted by Nugraha, where the accuracy value reached 93%.

RESEARCH METHODS

Research Stages

Below is an overview of the author's framework of thought. The research process starts with field studies and literature, data collection, pre-processing, translation, data labeling, word weighting, classification, evaluation, and visualization. Figure 2 shows the steps taken.

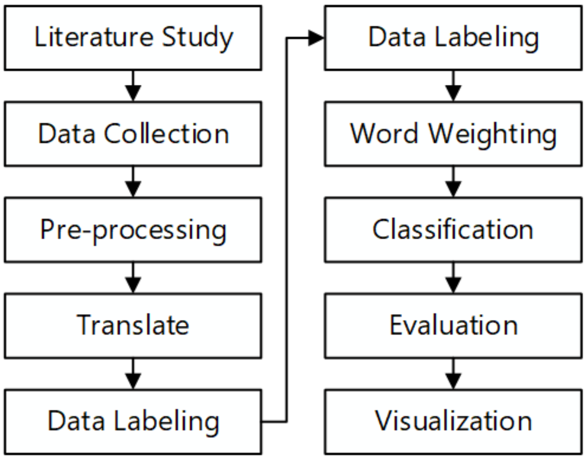

Figure 2 Frame of mind

A. Literature study

Literature studies are carried out to collect data or knowledge from several sources related to the research topic. In this literature, research is carried out by reading and searching journals related to research.

B. Data Collection (Crawling)

Crawling is a method of collecting information from the web. The crawl operates automatically, matching user-entered keywords with retrieved information(Eka Sembodo et al., 2016). in this research methodology is the collection of telemedicine application data on Twitter using the Rapid Miner tool to retrieve the data needed. By using Rapid Miner, one can sort data elements and save them into datasets. In this research, the author took the tweet data of the Telemedicine application contained on Twitter by filling out queries with the query "Telemedicine Application and Health Service Application," result type, country code, and limit on the Rapid Minner application. Then, the data that has been retrieved will be saved in CSV form.

C. Pre-processing

Pre-processing is done to create a dataset prepared for analysis. There are three stages tried in the pre-processing session of this research, namely.

1. Transformation

a. Cleansing

Data cleansing is the process of removing noise from data, including reading characteristics and other useless characters.

b. Case folding

Make all comments lowercase. A word consists of letters in several cases, including capital and lowercase letters. All letters are converted to lowercase to standardize upper and lowercase letters.

2. Tokenization

Breaking the reading or separating words can be called hyphenating sentences into words.

3. Filtering

Filter or eliminate useless terms at a later stage

a. Stopwords Removal

is to omit conjunctions such as "and," "which," "in," and so on.

b. Stemming

function to replace affixes with root words. Stemming is the process of switching between two languages. The Python Literati package can be used to do stemming in Indonesian.

D. Translate

The next step is data from tweets translated into English from Indonesia. The translation session is necessary because the next step will use the Vader Sentiment Library, which analyzes sentiment in English. Because vader sentiment is based on the English corpus. (Taufiq Anwar et al., 2023).

E. Data Labeling

Then, at the stage of labeling data with lexicon based using Vader lexicon. A lexicon-based feature is simply words that have been added to a dictionary or other reference material. Any word that contains positive or negative sentiment is subject to heavy processing. The purpose of a lexicon-based characteristic is to facilitate the understanding of the orientation of a single word.(Kurniawan & Adinugroho, 2019) Based on the conclusions of this study, the Vader Sentiment Library will be used to categorize sentiments and assess polarity. Labeling data by adding a new column containing the score and aggregated value of the Tweet data. A polarity score of 0 indicates neutrality, <0 indicates negative feelings, and >0 indicates a positive opinion.

F. Word weighting

After the next stage of data labeling, the weighting process is carried out. The term inverse document frequency (TFIDF) technique is used in the feature selection. The statistical term weighting technique, the TF-IDF approach, is often used as vectorization in analytical texts. The approach used by TF-IDF is to search comments for specific or pertinent terms—the feature selection procedure results in the token or short phrase, along with its TF-IDF weight.

G. Classification

At this stage, researchers will categorize the data using the Naive Bayes approach, which uses probabilistic calculations.
According to (Rachman et al., 2021). Naive Bayes is a simple classification technique that can calculate all occasions by adding the number of combinations and frequency of data points from the collected database. The categorization of positive and negative classes of the previous computational process results from analyzing the probability for each word in the document. i.e., the weight of the TF-IDF(Chatrina Siregar et al., 2020)

H. Evaluation

The Confusion Matrix is used to assess the accuracy of sentiment categorization models. The dataset is divided into training sets and tests before modeling can begin. The training aims to make learning data-driven and establish a pattern of whether comments made in class are positive or negative. Data testing is performed to evaluate the accuracy of the model. The accuracy of modeling sentiment analysis of telemedicine applications on Twitter using lexicon-based classifiers and naïve Bayes will be known based on test results with test data, which will produce a fusion matrix of True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN).

I. Visualization

The visualization phase is done after each step and process has been completed. Utilizing the Matplotlib package, the visualization phase of this research is complete. The result of this visualization is a pie chart image that also shows the sentiment class of each document, resulting in polarity. Wordcloud graphs show frequently used terms in each sentiment.

RESULTS AND DISCUSSION

The dataset used in this study came from Twitter. Researchers crawled the data using the rapid manner application. The data was taken from an account tweeted about telemedicine applications in Indonesia. The keywords used are telemedicine and health service applications and data withdrawal from April 1, 2023, to June 8, 2023.

1. Data crawling

Crawling is a method of collecting information from the web. The crawl operates automatically, matching user-entered keywords with retrieved information (Eka Sembodo et al., 2016). The process is to Visit every document on the website, starting from listing all the website URLs and tracking them one by one. Then, the data will be collected. In this case study, the data used was provided from the Twitter API(Dikiyanti et al., 2021). Crawling tweet data is done by input queries, searching on rapid miner applications, and taking tweets mentioning user accounts by writing keywords for telemedicine applications and health care applications. The data Crawling Process is shown in Figure 3.

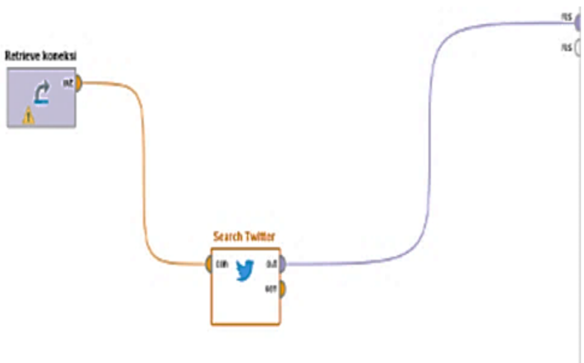

Figure 3. Data Crawling Process

2. Pre-processing

This process is done to make as many non-standard languages as possible into standard languages and to clean the data from noise so that the data becomes more structured and is ready for analysis. (Laurensz et al., 2021)Text pre-processing divides the text into smaller units (tokens) and treats text before text analysis. Tokenization, normalization, and filtering are used before the textbook is processed.

a. Transformation

1) Case folding

It plays a role in replacing all text with lowercase.

Table 1 Case Folding Results

Before	After
Bener bgt nih tulisannya mbak @armelzahfauzi semakin banyak aplikasi kesehatan hari gini.\n\nLayanan Telemedicine, Naik Pamor di Era Pandemi https://t.co/nwCBUCDtTO	bener bgt nih tulisannya mbak @armelzahfauzi semakin banyak aplikasi kesehatan hari gini.\n\nlayanan telemedicine, naik pamor di era pandemi https://t.co/nwcbucdtto
Telemedicine \nTelemedicine merupakan layanan aplikasi telekomunikasi untuk mendapatkan informasi dan pelayanan medis jarak jauh. Selama pandemi ini, penggunaan aplikasi telemedicine melejit terutama untuk layanan konsultasi kesehatan jarak jauh https://t.co/DEKXPnphpd	telemedicine \ntelemedicine merupakan layanan aplikasi telekomunikasi untuk mendapatkan informasi dan pelayanan medis jarak jauh. selama pandemi ini, penggunaan aplikasi telemedicine melejit terutama untuk layanan konsultasi kesehatan jarak jauh https://t.co/dekxpnphpd

2) Remove URL

To remove the URL from the text. Use remove URL in see as in table 2.

Table 2. Hasil *Remove* URL

Before	After
bener bgt nih tulisannya mbak @armelzahfauzi semakin banyak aplikasi kesehatan hari gini.\n\nlayanan telemedicine, naik pamor di era pandemi https://t.co/nwcbucdtto	bener bgt nih tulisannya mbak semakin banyak aplikasi kesehatan hari gini layanan telemedicine naik pamor di era pandemi

b. Tokenizing

Tokenizing reduces readings to single words and eliminates reading characteristics such as numbers. Usually, when composing tweets or tweets, account owners only use general terms, sometimes even slang. Tokenization replaces these standard terms with easier-to-understand alternatives. The results of tokenizing can be seen in Table 3 below.

Table 3. Tokenizing Results

Before	After
menggunakan platform digital untuk kemudian diantarkan ke rumah lawan pandemi	['menggunakan', 'platform', 'digital', 'untuk', 'kemudian', 'diantarkan', 'ke', 'rumah', 'lawan', 'pandemi']
dalam kasus telemedicine spt ini seharusnya ada regulasi khusus yang mengatur dokternya lagi lagi aplikasi hanya media saja	['dalam', 'kasus', 'telemedicine', 'spt', 'ini', 'seharusnya', 'ada', 'regulasi', 'khusus', 'yang', 'mengatur', 'dokternya', 'lagi', 'lagi', 'aplikasi', 'hanya', 'media', 'saja']

Filtering

Filtering is the process of deleting or saving word options. Here is the procedure to remove words and symbols from the data that will not be included in the sentiment analysis step later.

StopWord Removal

By setting the language you want to filter, you can try removing the stopword from the text (for example, and, or, this...). Stopwords are distributed through NLTK servers and can be downloaded for linguistic reasons. Table 4 below displays stopword results.

Table 4 Stopword Removal Results

Before	After
beli obat di halodoc trnyata cpt jg nyampenya mana ongkirnya murah cuma rb dri bali	buy medicine at halodoc, it turns out it's fast, where the postage is cheap, only rb from bali
kebetulan gue lagi isoman karena positif covid asuransi swasta ngasih benefit jadi enak aja ngurusnya kalau menurutku memang bpjs bagus tuh kerjasama dengan telemedicine kayaknya bakal terwujud di aplikasi fitaja setahuku sekarang masih proses pendirian	Incidentally, I'm soman because I'm positive for Covid. Private insurance gives benefits, so taking care of it is easy. I think BPJS is good. The collaboration with telemedicine seems to be realized in the Fitaja application. As far as I know, it's still being established.

Stemming

Stemming serves to replace affixes with root words. To carry out stemming, Indonesians can use the Python Sastrawi library. The results of stemming can be observed in Table 5.

Table 5 Stemming Results

Before	After
bawa ke vet atau tanya halodoc yaaa supaya muntahnya bs berhenti	['bawa', 'vet', 'halodoc', 'yaaa', 'muntahnya', 'bisa', 'berhenti']
halodoc jadi aplikasi telemedicine paling banyak dipakai orang indonesia	['halodoc', 'aplikasi', 'telemedicine', 'dipakai', 'orang', 'indonesia']

3. Translate

The next step after pre-processing is to convert tweet data from Indonesian to English. The translation phase has ended, and the next phase will use the English-language Vader Sentiment Library. The translation results are displayed in Table 6.

Table 6 Translate results.

Before	After
terima kasih atas fasilitas telemedicine dan paket obat isoman gratisnya walaupun paket obat yg datang tidak sesuai dgn yg disampaikan	['terima', 'kasih', 'fasilitas', 'telemedicine', 'paket', 'obat', 'isoman', 'gratis', 'paket', 'obat', 'yg', 'sesuai','tidak, 'dgn', 'yg','sampai']
jujur gue gatau itu knp nder klo malu mending pake aplikasi telemedicine kayak haldoc aja kan ga ketemu dokternya tuh cuma via chat krn yg paham gini cuma dokter	['jujur', 'gue', 'gatau', 'knp', 'nder', 'klo', 'malu', 'mending', 'pake', 'aplikasi', 'telemedicine', 'kayak', 'haldoc', 'aja', 'ga', 'ketemu', 'dokter', 'tuh', 'via', 'chat', 'krn', 'yg', 'paham', 'gini', 'dokter']

4. Data Labeling

The previous process will be labeling Tweet data. Data collection uses a language model based on language as the primary language. A key finding of the study is that using the Sentiment Vader library is the best way to classify sentiments and assess polarity. They label data by creating a new column with labels from Twitter data with a polarity score of <0 representing negative sentiment, =0 for neutral sentiment, and a polarity score of >0 illustrating positive sentiment. The labeling results can be seen in Figure 7 below.

Table 1 Labeling Result

Data Tweet	compound_score	Sentiments
saat ini kita memasuki era telemedicinese buah terobosan tekno dlm dunia kesehatan aplikasi health kian populer dgn biaya yg murah	0,4215	positive
aplikasi telemedicine yg sdh dicover namanya apa ya min	0	Netral
hari ini ke rs lagi penyebabnya pemberian obat lameson buat sebum yg dikira dokternya saya autoimmune gw mulai skeptis sama telemedicine	-0,2263	Negatif

The number of pre-processing and labeling data from sentiment documents is 1248 tweets, with 908 positive tweet data, 168 neutral tweet data, and 172 negative tweet data. The labeling results can be seen in Figure 4 below:

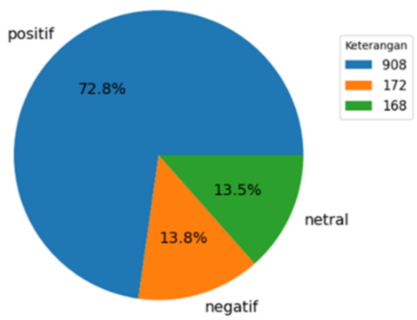

Figure 4. Hail Stage Labeling Vader Sentiment

JURNAL RISET INFORMATIKA **P-ISSN: 2656-1743 |E-ISSN: 2656-1735**
Vol. 5, No. 4. September 2023 DOI: https://doi.org/10.34288/jri.v5i4.244
Accredited rank 4 (SINTA 4), excerpts from the decision of the DITJEN DIKTIRISTEK No. 230/E/KPT/2023

Term Frequency-Inverse Document Frequency (TF-IDF)

Next, the term weighting process is carried out, which will be given a weight, where the weight proves the importance of the term to the document. Next, the calculation of weights on the terms sought in each document aims to recognize the usefulness and similarity of times in the paper. The more words that arise in the document, the greater the value or weight of the word. The calculation of term weighting begins by finding the term frequency (occurrence of words) in each document, as shown in Figure 8 below.

Table 8 TF-IDF Weighting Stage Results

Word	W		
	D1	D2	D3
aplikasi	0,22185	0,22185	0,22185
telemedicine	0,22185	0,22185	0,22185
semakin	1	0	0
banyak	0,52288	0	0
mudah	1	0	0
akses	1	0	0
penting	0	1	0
pasien	0	1	0
isoman	0	1	0
gratis	0	1	0
memang	0	0	1
krusial	0	0	1
dengan	0	0	0
layanan	0	0	0
semua	0	0	0

Classification

Classification using the Naive Bayes Classifier algorithm, which aims to be able to learn data patterns, especially sentiment in training data, so that it can make decisions for test data that will be inputted through creating documents in Python programs and generate sentiment for the test data, At this stage the dataset is divided into two parts are training data and testing data with 3 test and training experiments, namely 90% data train and 10% test data, 80% train and 20% test data and 70% train and 30% test data. Sourced from research.(Gormantara, 2020).

Table 9. Train & Testing Data Comparison

Data Train	Data Test
90%	10%
80%	20%
70%	30%

In the naïve Bayes method, several test data division scenarios are used, as in Table 10. Data train and subsequent datasets are processed with the naïve Bayes classifier method. The results of dataset sharing accuracy is 90%; 10% is 76%. For 80% data sharing, 20%, the accuracy is 75%, and for 70%, 30%, the accuracy is 74%. The best accuracy results are at a dataset ratio of 90%:10%, i.e., 76%.

Table 11 Training & Testing Data Comparison Results

Data train : Data test	accuracy	Precision	Recall
90% : 10%	0,768 (77%)	0,697 (70%)	0,768 (77%)
80% : 20%	0.748 (75%)	0.712 (71%)	0.748 (75%)
70% : 30%	0,741 (74%)	0,689 (67%)	0,741 (74%)

Evaluation

The evaluation aims to check the correctness of the classification results. At this stage, the evaluation results use the Confusion Matrix to determine the results of accuracy, precision, and recall. The results of the matrix confusion evaluation process with 80% training data and 20% testing data can be seen in Image 5 below.

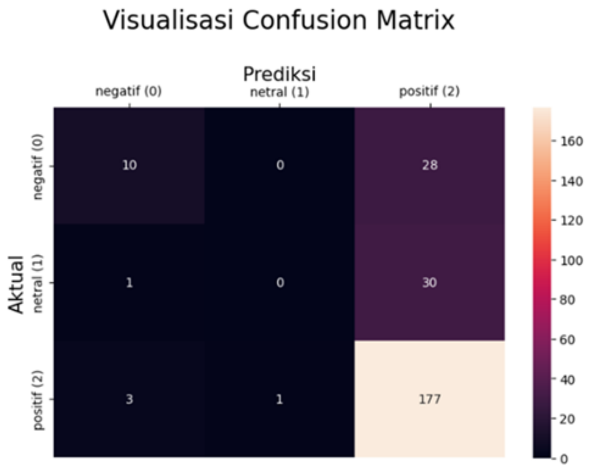

Figure 5 Matrix confusion

Table 11 Confusion Matrix Calculation Results

Accuracy	Precision	Recall
74%	98%	86%

Visualization

At this stage, researchers use the matplotlib library from Python to visualize the data processing results.

Piechart

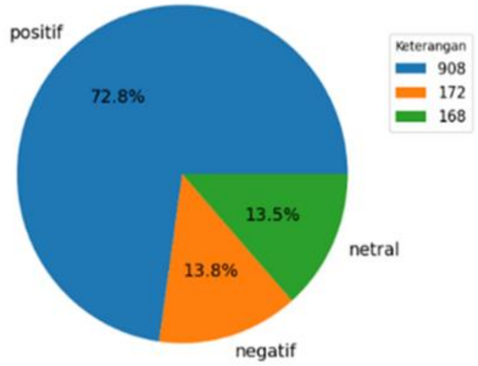

Figure 6 Sentiment Class Piechart

Based on Figure 6, 1248 tweet data have gone through the pre-processing stage. Tweet data is divided into three categories: positive, negative, and neutral. The positive style dominates with 908 tweet data, 172 negative, and 168 tweet data. So, it can be concluded on social media, Twitter, many positive tweets against telemedicine applications.

Wordcloud

Figure 7 Word Cloud Neutral Class

Based on Figure 7, The neutral word cloud above contains words that often appear in neutral class sentiment documents, namely "telemedicine application," "Halodoc," "consultation," "info," "check," "use," "online," etc.

Figure 8 Wordcloud Positive Classes

Based on Figure 8 Positive Wordcloud above, some words often appear in positive class sentiment documents, namely "telemedicine application," "Halodoc," "consultation," "service," "free," "thanks," and others. - other

Figure 9 Wordcloud Negative Class

Based on Figure 9. In the negative word cloud above, some words often appear in negative class sentiment documents, namely "telemedicine application," "Halodoc," "consultation," "sick," "no," "wrong," and "waiting." "and others.

CONCLUSIONS AND SUGGESTIONS

Conclusion

From the results of this study, it can be concluded that this study implements the use of Vader libraries to determine polarity. This research analyzes public opinion sentiment towards the telemedicine application on Twitter social media using lexicon-based and naïve Bayes classifier methods. The crawling data obtained was 7168 tweet data and, after the pre-processing stage, produced 1248 tweet data. Then, the tweet data is translated into English and divided into three classes, positive, neutral, and negative, using the lexicon-based method. Based on the labeling results using lexicon-based, 908 positive tweet data, 172 negative tweet data, and 168 neutral tweet data were obtained, so it can be concluded that the sentiment regarding public opinion towards telemedicine applications is primarily positive. Furthermore, modeling was carried out using the naïve Bayes classifier method. The performance accuracy of the naïve Bayes classifier method resulted in an accuracy of 74%, precision of 98%, and recall of 86%.

Suggestion

Based on the research results, researchers have made several suggestions that become input and consideration for further research. Since the dataset in this study only used Indonesian, future studies will compare tweet sentiment based on language and distribution of user location and use social media other than Twitter as a source of data collection. It is recommended that future studies use libraries other than the Sentiment Vader library, use more datasets, and use some classification techniques in addition to naïve Bayes classifiers.

REFERENCES

Annur. (2022). *Layanan Telemedicine yang Paling Banyak Digunakan di Indonesia, Apa Saja?* Databoks. https://databoks.katadata.co.id/datapublish/2022/04/07/layanan-telemedicine-yang-paling-banyak-digunakan-di-indonesia-apa-saja

Chatrina Siregar, N., Ruli, R., Siregar, A., Yoga, ; M, & Sudirman, D. (2020). Implementasi Metode Naive Bayes Classifier (NBC) Pada Komentar Warga Sekolah Mengenai Pelaksanaan Pembelajaran Jarak Jauh (PJJ). In *Jurnal Teknologia Aliansi Perguruan Tinggi (APERTI) BUMN* (Vol. 3, Issue 1).

Cikania, R. N. (2021). Implementasi algoritma naive bayes classifier dan support machine pada klasifikasi sentimen review layanan telemedicine halodoc. *Jambura Journal of Probability and Statistics*, *2*(2), 96–104. https://doi.org/10.34312/jjps.v2i2.11364

Dikiyanti, T. D., Rukmi, A. M., & Irawan, M. I. (2021). Sentiment analysis and topic modeling of BPJS Kesehatan based on twitter crawling data using Indonesian Sentiment Lexicon and Latent Dirichlet Allocation algorithm. *Journal of Physics: Conference Series*, *1821*(1). https://doi.org/10.1088/1742-6596/1821/1/012054

Eka Sembodo, J., Budi Setiawan, E., & Abdurahman Baizal, Z. (2016). *Data Crawling Otomatis pada Twitter*. 11–16. https://doi.org/10.21108/indosc.2016.111

Gormantara, A. (2020). *Analisis Sentimen Terhadap New Normal Era di Indonesia pada Twitter Menggunakan Metode Support Vector Machine*. https://www.researchgate.net/publication/342986951

Kemkes. (2021). *Aplikasi Telemedicine Berpotensi Merevolusi Pelayanan Kesehatan di Indonesia*. Kemkes. https://www.balaibaturaja.litbang.kemkes.go.id/read-aplikasi-telemedicine-berpotensi-merevolusi-pelayanan-kesehatan-di-indonesia

Kurniawan, A., & Adinugroho, S. (2019). *Analisis Sentimen Opini Film Menggunakan Metode Naïve Bayes dan Lexicon Based Features* (Vol. 3, Issue 9). http://j-ptiik.ub.ac.id

Laurensz, B., Sentimen, A., & Sediyono, E. (2021). Analisis Sentimen Masyarakat terhadap Tindakan Vaksinasi dalam Upaya Mengatasi Pandemi Covid-19 (Analysis of Public Sentiment on Vaccination in Efforts to Overcome the Covid-19 Pandemic). In *Jurnal Nasional Teknik Elektro dan Teknologi Informasi |* (Vol. 10, Issue 2).

Mahendrajaya, R., Buntoro, G. A., & Setyawan, M. B. (2019). *url : http://studentjournal.umpo.ac.id/index.php/komputek* Analisis sentimen pengguna gopay menggunakan metode lexicon based dan suppoort vector machine. http://studentjournal.umpo.ac.id/index.php/komputek

Rachman, R., Handayani, R. N., & Artikel, I. (2021). Klasifikasi Algoritma Naive Bayes Dalam Memprediksi Tingkat Kelancaran Pembayaran Sewa Teras UMKM. *JURNAL INFORMATIKA*, *8*(2). http://ejournal.bsi.ac.id/ejurnal/index.php/ji

Riyanto, A. (2021). Faktor-Faktor yang Mempengaruhi Pelaksanaan Telemedicine (Systematic Review). *Jurnal Manajemen Informasi Kesehatan Indonesia*, *9*(2), 174. https://doi.org/10.33560/jmiki.v9i2.337

Salehudin Basryah, E., Erfina, A., & Warman, C. (2021). analisis sentimen aplikasi dompet digital di era 4.0 pada masa pademi covid 19 di play store menggunakan algoritma naive bayes classifier.

Taufiq Anwar, M., Riandhita Arief Permana, D., STMI Jakarta, P., Sistem Informasi Industri Otomotif, P., Letjen Suprapto No, J., & Pusat, J. (2023). *Analisis Sentimen Masyarakat Indonesia Terhadap Produk Kendaraan Listrik*

Menggunakan VADER. *10*(1), 783–792. http://jurnal.mdp.ac.id

Yuniar, E., Utsalinah, D. S., & Wahyuningsih, D. (2022). Implementasi Scrapping Data Untuk Sentiment Analysis Pengguna Dompet Digital dengan Menggunakan Algoritma Machine Learning. *Jurnal Janitra Informatika Dan Sistem Informasi*, *2*(1), 35–42. https://doi.org/10.25008/janitra.v2i1.145